(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 January 2004 (22.01.2004)

PCT

(10) International Publication Number
WO 2004/008151 A2

(51) International Patent Classification[7]: G01N 33/68

(21) International Application Number:
PCT/HU2003/000056

(22) International Filing Date: 9 July 2003 (09.07.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
P 0202212 10 July 2002 (10.07.2002) HU

(71) Applicant *(for all designated States except US)*: **COMGENEX RT.** [HU/HU]; Bem Rpt. 33-34., H-1027 Budapest (HU).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **DARVAS, Ferenc** [HU/HU]; Zápor U. 11/B., H-1034 Budapest (HU). **ÜRGE, László** [HU/HU]; Kerékgyártó u. 3, H-1147 Budapest (HU). **DORMÁN, György** [HU/HU]; Kondorosi u. 80, H-1119 Budapest (HU). **KRAJCSI, Péter** [HU/HU]; Batyu u. 7, H-1025 Budapest (HU). **SZALAY, Dániel** [HU/HU]; Székely Bertalan U. 5, H-1062 Budapest (HU). **BÉRES, Mariann** [HU/HU]; Szépvölgyi U. 212., H-1025 Budapest (HU). **BUCSAI, Ágota** [HU/HU]; Kerék U. 26. IX/52, H-1035 Budapest (HU). **NAGY, Tamás** [HU/HU]; Hunyad Lejt 28, H-1121 Budapest (HU). **KÁLMÁN, Ferenc** [HU/HU]; Sáfrány U. 40 VIII/83, H-1116 Budapest (HU).

(74) Agent: **KORMOS, Ágnes**; Váci Út 66., H-1132 Budapest (HU).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 2004/008151 A2

(54) Title: NEW COMBINATORIAL PROTEIN MARKER COMPOUND LIBRARIES AND METHODS FOR THEIR PREPARATION AND UTILIZATION

(57) Abstract: The invention consists substantially of combinatorial, protein-marker molecule libraries, having groups that may be chemically and photochemically activated, with or without reporter groups, connected in different diversity points or spatial arrangements via side chains around a common molecular core. The invention consists furthermore of combinatorial libraries of marker units having chemically or photochemically reactive marker groups, reporter groups, and various other types of side chains attached variably to a common molecular, preferably lysine-based, structural core. The invention consists furthermore of a method for combinatorial chemical tethering, where a side chain having a terminal functional group, preferably an amino group, and to which a marker group is optionally attached, is introduced in optimal position ensuring structural diversity and protein binding effectiveness. The invention contributes to simplifying complexity of proteome in different tissues and to the identification of protein markers for diagnosis by alternating different labeling pattern in healthy and diseased tissues.

NEW COMBINATORIAL PROTEIN MARKER COMPOUND LIBRARIES AND METHODS FOR THEIR PREPARATION AND UTILIZATION

The invention pertains to substantially new combinatorial protein marker compound libraries including methods for their preparation and utilization.

As a result of the mapping of the human genome, thousands of new proteins remain to be identified. Identifying these proteins as potential drug targets will constitute one of the most important challenges for drug research in coming years. Numerous methods have been developed that associate new proteins discovered by means of the genome with particular diseases, including comparative 2D electrophoresis (A. Görg, Proteomics, 2000, July 3) and isotope labeling combined with mass spectrometry (S.P. Gygi et al, Proteomics, 2000, July 31).

In many cases, these methods do not yield results when, for example, proteins are available only in low concentrations. Separating, detecting, and identifying large numbers of proteins is also difficult. The most significant demand on newly emerging technological solutions is that they identify not only the protein related to a particular disease, but also the small molecules, which are capable of influencing these proteins.

The binding of a small molecule validates the target protein, and inversely the binding to the target confirms the biological activity of the ligand (small molecule). Supplementary expression tests can also be used to demonstrate not only the function of the protein, but also whether the disease state is affected by the binding. (G. Dormán, et al. Current Drug Discovery, 2001, 1, 21-24) The use of known and novel ligands offers, numerous possibilities for studying and characterizing both known and unidentified proteins, as shown in the following chart.

Ligand/ Substrate \ Protein	Known target	Potential target
Known drugs	Positive control validation Known interaction New interaction Side effects	Assigns new function for known protein Assigns function for new protein New target identification New mode of action
Drug candidates	New hit discovery Confirmation of drugability	New interaction discovery Confirmation of drugability Potential therapeutic relevance

With the development of combinatorial chemistry, in which every possible combination of building blocks are attached to a central core, it is now possible to produce a large number of new small molecules, as well as new analogues of previously known effective drug molecules.

Recognition of this fact has allowed for the effective high-throughput identification and classification of new and known proteins based on the affinity-based interaction of a large number of small molecules as per the matrix in the above chart. The rapid advances in bioanalytics and bioinformatics have also supported this possibility.

Since the affinity-based methods identify and functionally validate the proteins encoded by newly discovered genes, primarily using synthetically produced small organic molecules, this new approach is called chemical genomics and proteomics.

The affinity-based methods have also been employed for isolating and identifying proteins.

In affinity chromatography, small molecules are bound to cellulose or other polymers, and a protein mixture is streamed through this so-called affinity column.

The proteins leave the column separately and at different speeds depending on the binding strength to the small molecules.

Another affinity-based method, called affinity labeling, involves a specific chemical or photochemical reaction occurring by means of a reactive marker group, resulting in the formation of a stable covalent bond between the protein and ligand. The main advantage afforded by this method is that the adduct remains stable even after denaturation of the protein, so that the ligand and the reporter groups linked to it remain in a state of attachment to the protein (at the binding site).

Basically, this actually forms the basis for the affinity labeling technique, and, as long as the ligand contains a reporter group, it introduces a label to the target by the formation of a covalent adduct. The so-called "reporter" structural unit or group may be a fluorescent, radioactive, biotin, spin, or other unit used for the purposes of marking the target. (The name "reporter" refers to the fact that the group provides information regarding conformation, binding, etc.)

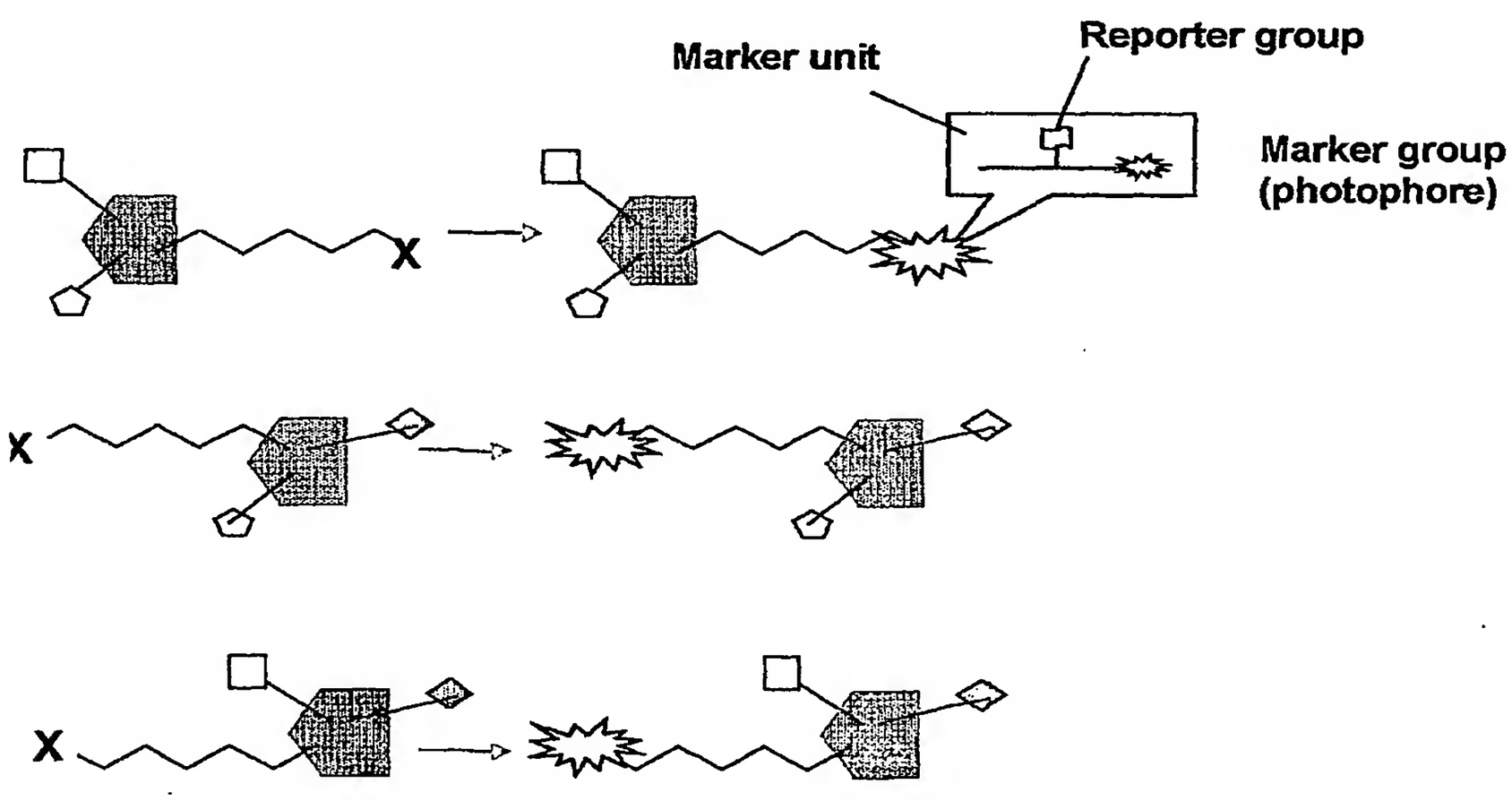
When groups that may be chemically activated (for example aziridin or epoxid groups) are used, the ligand or substrate is often used up in non-specific reactions with the nucleophiles before binding.

The wide use of photochemically activated groups (photophores such as benzophenone or aromatic azides) is attributable to the fact that such groups are remotely controllable 'clean reagents' and that they have numerous other favorable properties, in contrast to groups that are simply chemically reactive. Thus, they are stable in the "dark" and in a biological non-covalently bound state. A photochemical reaction resulting in a covalent bond in a well-defined time will occur only upon irradiation of light by the researcher. (Dormán G., Fénnyel aktiválható biológiailag hatékony vegyületek (Photo-activatable biologically efficient compounds.), Kémia Újabb Eredményei (The Newest Results of Chemistry), Volume 89, Akadémiai Kiadó, 2001, in Hungarian). It is important that the wavelength of light sufficient for excitation should not harm the protein (>350 nm).

Normally, in designing a combinatorial library, use of groups that may be activated chemically and photochemically is avoided, for precisely the reason that such groups react easily with proteins during biological interaction.

Our invention is based on the recognition that in chemical genomics we may obtain a large amount of biological affinity-based information quickly using combinatorial libraries appropriately marked with chemical or photochemical affinity-labeling groups, a technique that may be used for determining of the binding profiles, quick classifying the target proteins rapidly, or directly identifying proteins and creating high-throughput biological screening systems.

Surprisingly, in producing combinatorial marker libraries, it was found that it was very advantageous if the chemically and photochemically reactive marker groups (such as the photophores) and reporter groups were attached to the library as a marker unit, in a single, parallel, robotized step, through functional groups initially introduced at different diversity points of the structure. The functional group may be positioned directly on the core structure, or preferably, as the terminal group of a side-chain what is called tether.



Based on above, this invention involves combinatorial protein marker molecule libraries containing groups that may be chemically or photochemically activated, with or without reporter groups, attached at different diversity points or spatial directions via side chains around a molecular core.

The invention consists furthermore of combinatorial marker unit libraries having chemically or photochemically reactive marker groups, reporter groups, and various types of side-chains attached variably to a common molecular, preferably lysine-based, structural core.

The library described as part of this invention contains: benzophenonee and nitro-phenilazide groups as marker groups; biotin and fluorescent groups as reporter groups; saturated carbon chains as side chains; and polyethylene-glycol units.

The invention consists furthermore of a method for combinatorial chemical tethering, whereby a side chain is introduced in optimal position, ensuring structural diversity and protein binding effectiveness. Preferably, a marker unit is attached to a side chain with a terminal functional group, preferably an amino group.

The invention consists furthermore of the application of tethered combinatorial libraries toward the study of non-covalent interactions between affinity-based biopolymers, preferably proteins and small molecules, such as ligands, substrates and other compounds, either directly, or immobilized to a solid support, preferably using affinity chromatography, chemical microarrays, or microchips.

The tethered chemicals immobilized to the solid support described as part of this invention may be applied toward the study of interactions between macromolecules, preferably proteins and small molecules, such that they are used in established protocols or processes for using proteins, DNA, or any other type of reader chip, that have been developed for application to the study of molecules immobilized to microchips or microarrays .

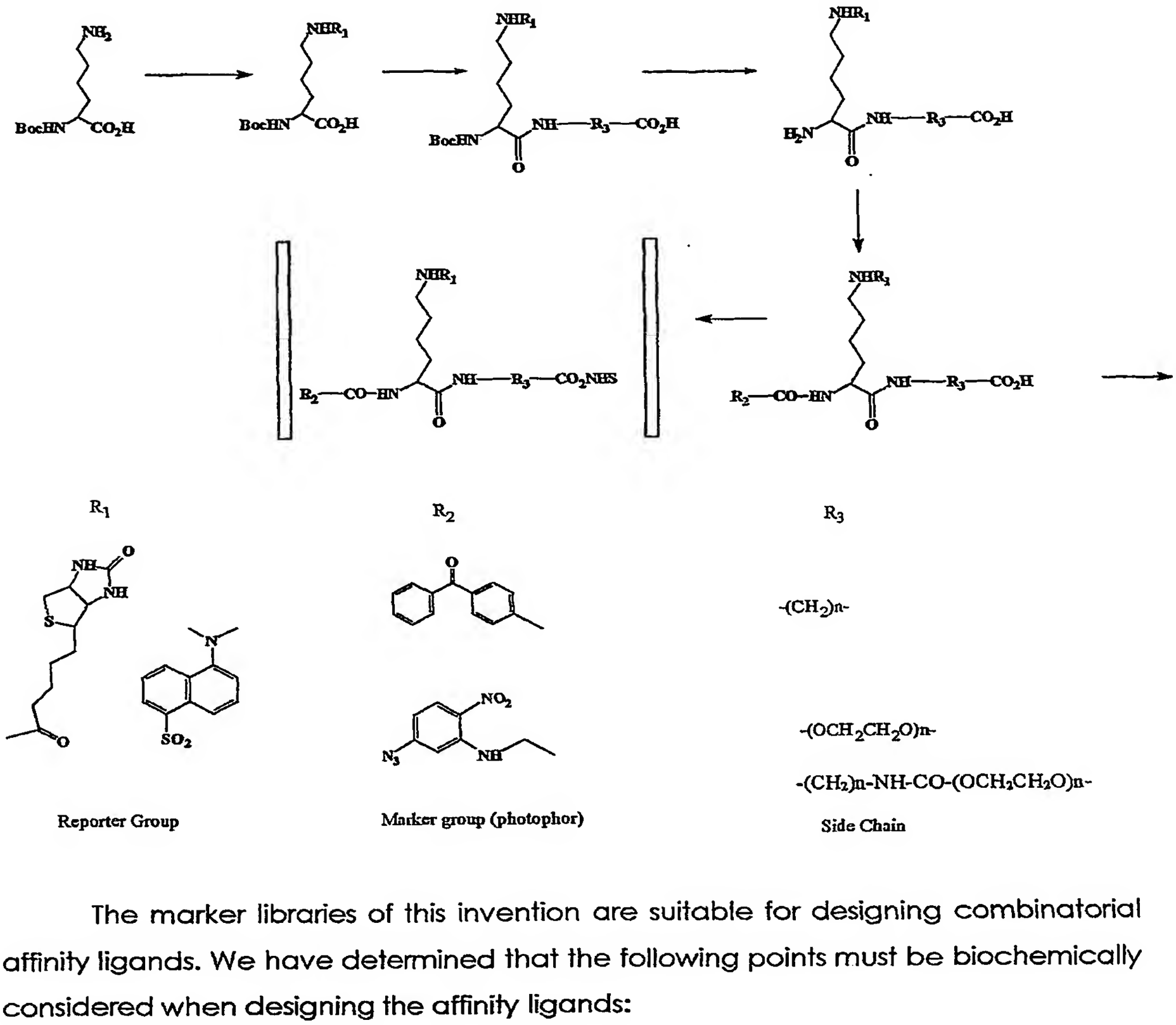
The invention also consists of a method for robotized parallel derivatization, in which the side chain or, directly, the marker unit is linked to the intermediate of the molecular library.

The invention consists furthermore of the application of a high throughput biological test that enables parallel affinity labeling of a large number of samples obtained from different tissues including the detection and separation of covalently bound proteins. This method is also suitable for identifying protein markers specific to diseases and may also be used as a diagnostic method. Application of the covalent labeling method contributes to simplifying complex proteomics.

Our invention consists furthermore of the application of a high-throughput analytical method suitable for sequencing the covalently bound proteins using mass spectrometric methods, and for comparing them to known sequence databases.

Based on our invention, a combinatorial library of simple marker units may also be formed that are attached en masse to appropriately prepared, small molecule libraries using parallel, robotized methods. The library of the marker unit is actually the set of all possible combinations of various types of attaching side-chains, (photo)reactive groups, and different reporter groups.

Preferably, this produces a lysine-based branched system, where the marker and reporter groups, as well as the quality of the side-chains, can be varied.



The marker libraries of this invention are suitable for designing combinatorial affinity ligands. We have determined that the following points must be biochemically considered when designing the affinity ligands:

1. The binding of the affinity-based probe molecules should expectably occur at the same site as in the case of ligands not containing these modifications, and their biological activation should be in the same order of magnitude.

2. The formation of the covalent bond and its irreversible activation or inactivation should be proportional to the intensity of the signal transferred by the reporter

group, in other words, besides the bound proteins, the nonspecific signals should be kept to a minimum.

3. In the case of photochemical activation, the excitation wavelength of the light employed should not cause any damage to the protein; said wavelength should not be less than 320 nm, while the value of □max (the molar extinction coefficient) should be high.

4. The covalent adduct should be stable under in the condition of chemical and enzymatic protein fragmentation. This is inherent in the specific mechanism of the chemically or photochemically reactive groups applied.

5. The covalent modification of the protein should preferably be point or regio-selective to some degree. This property provides a single, modified or labeled protein fragment with the labeling of one or two neighboring amino acids. This makes easier the identification of the protein using MS-based sequence identification or via comparison with sequence databases.

In the affinity labeling experiment that is part of the present invention, the tissue sample containing the target protein is incubated with the ligand for a short period of time. In the case of chemically activable groups, the crosslink forms directly. In the case of photoreactive groups, if the procedure is carried out in the "dark," the photochemical reaction characteristic of the given group takes place when the non-covalent linkage formed is irradiated by a wavelength appropriate for the photosensitive group, and an irreversible bond forms between the receptor and the ligand. This bond can be detected by means of the reporter group in several ways:

Since both the appropriate reporter group and the ligand are incorporated into the protein (in the binding region) the ligand binding protein/proteins can be distinguished from the polypeptide that does not show binding activity/affinity.

In addition to being suitable for detection, biological reporter groups are also suitable for separation: for example, a protein-ligand adduct containing a biotin unit

can be easily separated on an avidin affinity column. The separated proteins may be broken down into smaller fragments using enzymatic or chemical methods and subsequently analyzed using mass spectrometry or classical sequencing analysis. If the protein fragment containing the binding region is known, then without labeling, the exact location of the binding site can be identified directly by means of its MS fingerprint without recourse to special reporter groups.

The affinity-based interaction labeling experiment, using ligands made chemically or photochemically reactive by means of the present invention, provides information on the following levels:

- It identifies a protein-ligand interaction in a given cell-lysate or tissue sample, where the protein interacting with the ligand may have a known or unknown function, but its sequence should be identifiable using the DNA sequence database of the human genome,
- The region of the binding site within the protein can be identified,
- The amino acid sequence within the binding region of the protein may be identified from several perspectives, with the result that the minimum active fragment of the binding region can be determined,
- Under favorable conditions, the function of the protein and its role in chemical communication between and within cells can be identified, by irreversible activation and inactivation,
- It may be used to determine the expression level for a given protein in both diseased and healthy states,
- It may be used to characterize non-functional, transport processes, such as efflux pumps (MDR).
- Using differential experiments in healthy and disease tissues; disease marker proteins can be identified in diagnostics

In one embodiment of this invention, the reactive marker group/units are attached expectably to biologically active chemicals. It is known that in one group of affinity-labeling analogues (so-called exo-type tethered ligands; for naming, see: Baker: *Design of Active Site Directed Irreversible Enzyme Inhibitors*, 1967, John Wiley &

Sons, Inc., New York), the light-sensitive group is linked to the ligand via a side chain. This method is advantageous in isolating the receptor. In such cases, when the pharmacophore is a part of the reactive group and mimics its structural units, the mapping of the binding region of the receptor (endo-type) becomes possible. When exo-type affinity probes are used, the reactive group must be introduced at a sterically flexible site, far away from the pharmacophor, in order not to change the conformation of the compounds.

When synthesizing the reactive marker probes associated with this invention, the following points should be considered:

1. During the synthesis of the marker library elements (such as the linking of the marker unit to the tethered library) the chemically reactive group, or photophore, must not undergo degradation. For this reason, it is preferable that a stable precursor should be formed or built into the molecule to be converted into a photoreactive group during the last step of the reaction (the linear approach).

2. The reactive group or its precursor may be attached most simply directly to an available functional group of the natural ligand (endo-type), or we may form the photoreactive group by directly appending a simple group (semi-synthetic approach, such as when a substituted-benzophenonee is formed by directly benzoylating an aromatic ringbenzophenone).

3. In the absence of a suitable functional group, it is advantageous to introduce a nucleophile group to the molecule via a side-chain, with or without a "tether". In the last step, a synthon or a heterobifunctional reagent photophore containing the photophore is attached to this nucleophile group (the convergent approach.)

For our invention, the latter solution seems to be the most advantageous for the reasons of both synthetic parallelization and design.

Even in this situation, sterically large-sized reactive marker units (e.g. benzophenonee photophore) may exert a significant effect on the molecular conformation and biological activity. In this case, it is practical to attach a side-chain (a tether), occupying a small space, to the freely variable unit of the molecule containing the terminal group suitable for functionalization. The reactive or photoreactive groups may then be attached to the "tether," preferably with the reporter group, thus forming a separate marker unit. Another obvious advantage of the above "tether" ligand technique is that the bioactive ligand may be bound to a solid phase support via a side chain, making it possible to then pre-purify the binding protein using affinity chromatography.

When preparing the affinity-based marker libraries associated with this invention, the following should be considered:

Photoaffinity and other affinity-based bioconjugate techniques require careful planning and modeling if biological activity is to be sustained.

When planning the synthesis of tethered small molecule libraries, the most important task is to develop functional groups on the various diversity points of the molecular core, to which the marker unit may be attached, either directly or via another side-chain, preferably in the last step of synthesis. Thus, in planning the synthesis, those bifunctional or masked chemical reagents (building blocks) should be included in the standard reagent set, which results in these chemicals. Consequently, the production can be both parallelized and robotized.

In order to sustain biological activity, placement of the tether can be determined in 3 different ways, when analogue molecules are desired to be tested on a known protein:

Based on known affinity probe compounds or experience with affinity chromatography or QSAR data or 3D docking results reported in the literature, as long as the 3D structure of the protein is known.

The excited photophore has a greater chance of reaching a functional group if it attaches itself to a flexible alkyl side chain (also called a tether, see below).

In the discussion to follow, we analyze the role of the side chain used in the procedure specified by this invention:

The flexibility of the attaching side chain ('tether') substantially effects the success of affinity labeling. Selection of this side chain depends both on the set goal and on the characteristics of the target protein. The selection of a rigid or flexible linker generally means a compromise between two conflicting factors. More flexible and longer side chains increase the degree of freedom and offer a greater opportunity for covalent bonds to form, though they may result in a signal more distant from the binding region or in multiple points of attachment. If the "tether" is too short, the photoreactive group may embed itself deeply between the functional groups of the ligands, depending on ligand conformation, and may not be capable of effective signaling. Furthermore, intermolecular reactions may also occur. A rigid tether will probably label a single amino acid point-selectively, although its effectiveness is expected to be less, since the probability of forming the reactive protein parts is proportional to tether flexibility.

In summary, the flexibility and length of the "tether" depends mainly on the set goals. For identifying a new, unknown receptor protein (e.g. in the case of natural materials), a longer, more flexible side chain is more beneficial. In other instances, when the main goal is to spatially map the receptor, a rigid side chain of known length may provide the desired information. In the latter case, it is practical to test the photoreactive group on various positions of the bioactive ligand, so that the regio-specificity and effectiveness of the formation of the covalent bond is studied from various "points of view".

With the hydrophilicity or hydrophobicity of the flexibly binding side chain, the attaching photoreactive group may be guided towards various protein regions. For the labeling of protein segments embedded in the membrane, it is preferable to choose the hydrophobic „tether" (except in the case of ion channels), while for the intercellular segments of the receptor, applying the hydrophilic side chain is more advantageous. This also increases the general water solubility of the affinity ligand.

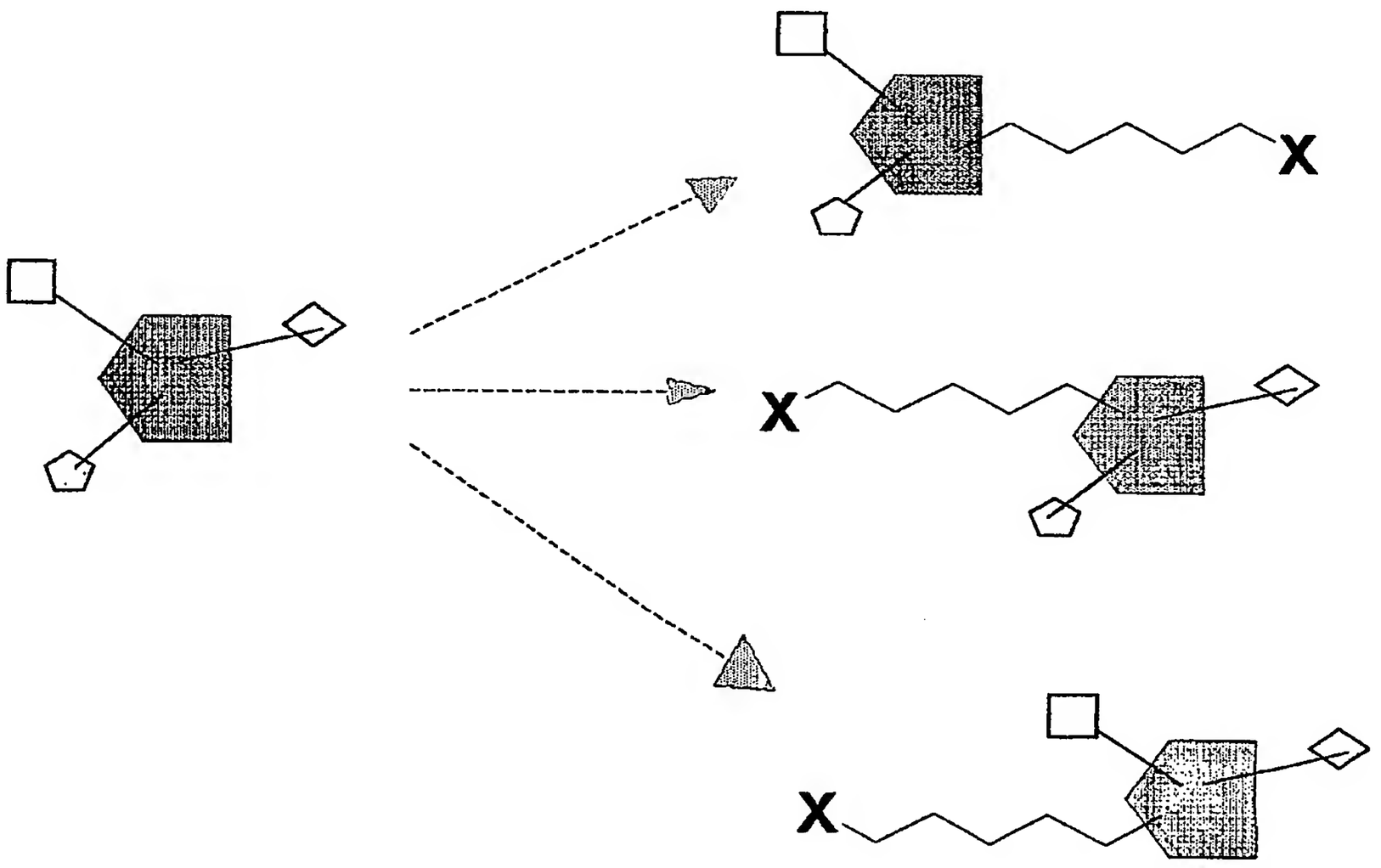
The main advantage of the solution proposed by this invention is that the tethered combinatorial library can be produced with high yield. Another important factor is that under robotized conditions, the functionalization and "tethering" of

known and effective synthesis of the ligand or library can be carried out using minimal modification.

In one advantageous variation of the invention, we produce marker groups using a combinatorial approach (marker library). Accordingly, for unknown proteins or small molecule families where we do not have ample amounts of information for the above, we apply a combinatorial approach. This results in placing the tether at various points around the molecular core, in those positions that are actually diversity elements of the molecule.

Using chemical or photomarker units attached to the new combinatorial libraries, the drugability of chemicals can be determined, and the protein target and attached biological ligands or substrates can be identified in a single step.

The combinatorial chemical libraries associated with this invention typically contain 3 or 4 diversity points, which generally point in different spatial directions.



The reactive marker groups can be placed in an appropriate, smaller, representative group of the library at variable points in such a way that they effect the expected biological activity of the unmodified chemicals only minimally.

The advantages of the invention are summarized as follows:

By enhancing affinity labeling experiments and detection with high-throughput and supplementing it with computerized data analysis, the high-capacity, parallel derivatization of the combinatorial libraries for producing derivatives resulting in covalent bonds (preferably with photoreactive groups) may be used as a method for fast protein profiling that allows for the detection and separation of the marked proteins, as well as their structure-based, sequential or functional characterization and classification. This method is also suitable for identifying protein markers specific to diseases and can be used as a diagnostic method. Use of the covalent labeling method contributes to simplifying complex proteomics.

This method may also be developed without reporter groups when previously identified, preferably recombinant proteins are used, through the application of mass spectrometry (MS), by seeking out MS photophore fingerprints. In this case, groups that may be activated (photo)chemically are necessary to be used, as these groups provide characteristic MS fragment patterns.

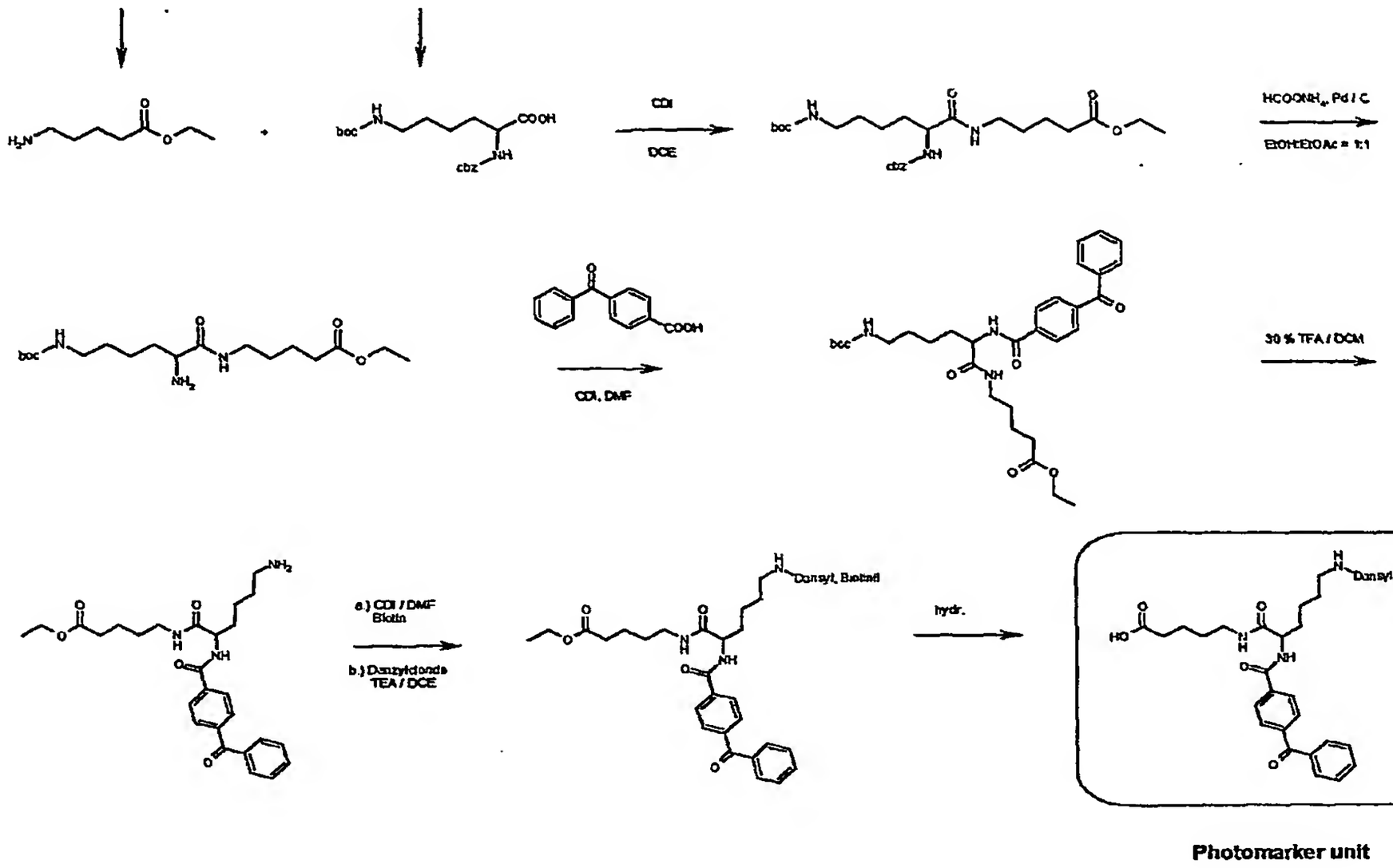
Our invention is introduced through the following examples, though it is to be understood that its range of applications is not restricted to these cases only.

Examples:

Example 1.

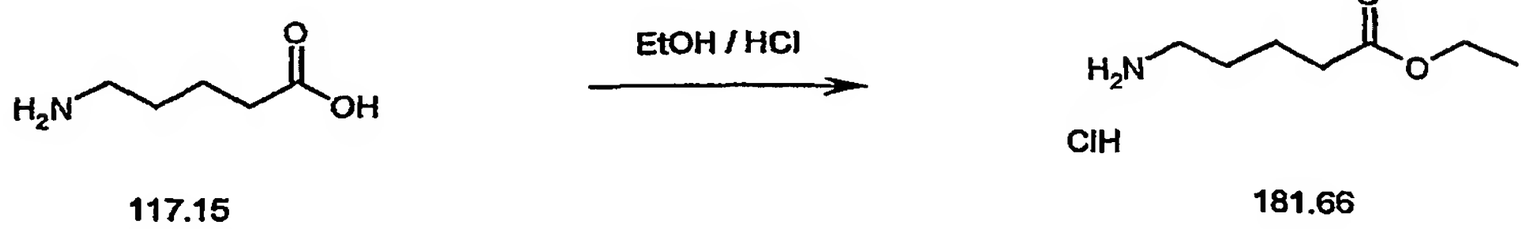
The synthesis of benzophenone-biotin and benzophenone-dansyl marker units

Schematic diagram of the reaction:



A./ Preparation of the initial intermediate:

5-amino-valeric acid ethyl ester hydrochloride:



117.15

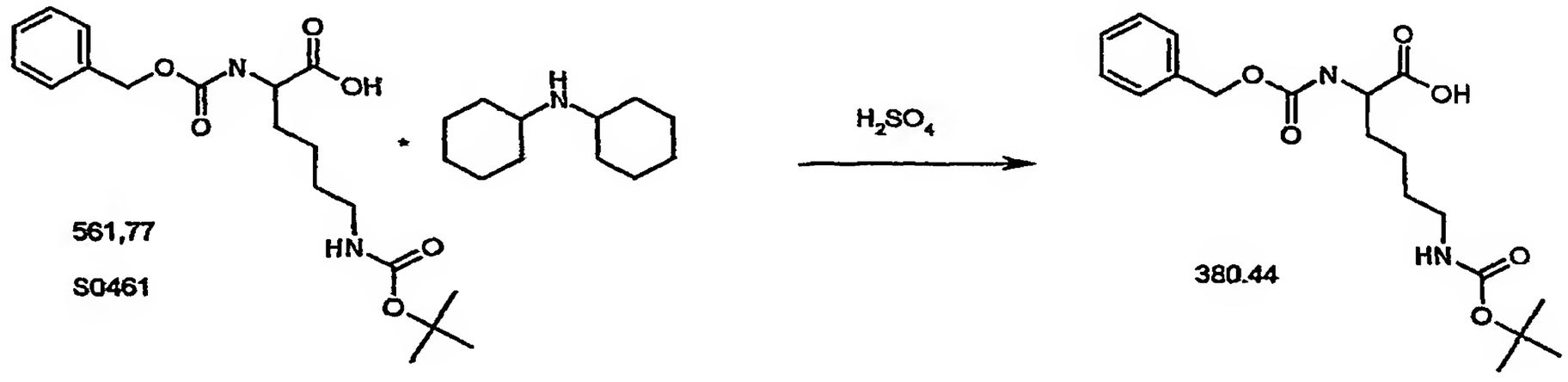
S1798

181.66

5.0 g (42.7 mmol) 5-amino-valeric acid was dissolved in a freshly prepared dry ethanolic-HCl solution (ca. 40 ml), and the solution then stirred at room temperature

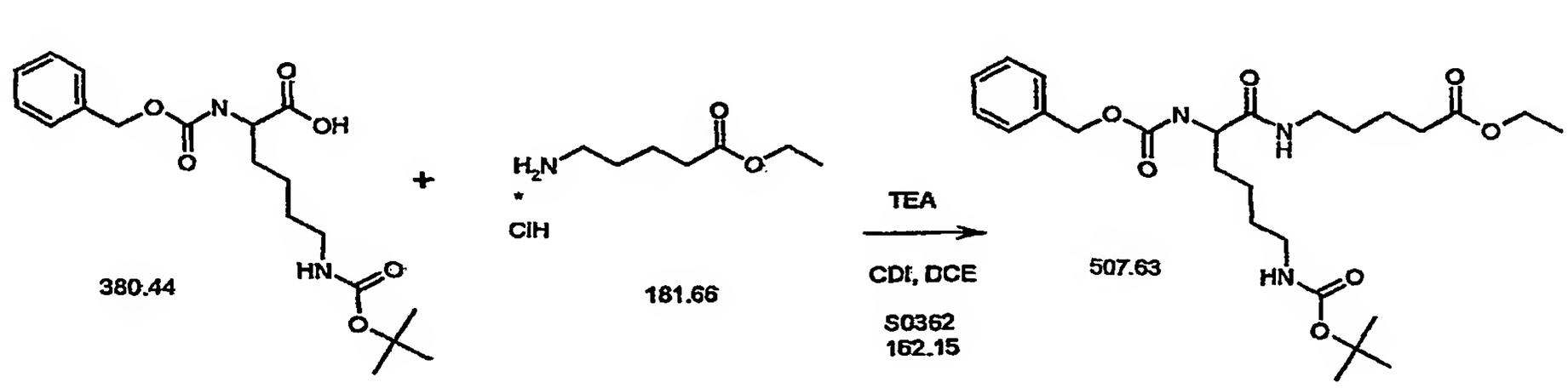
for 2-3 hours. TLC: $CHCl_3$: MeOH = 4:1, $R_f \approx 0.3$. When the initial acid had disappeared, the mixture was concentrated in a rotary evaporator under reduced pressure until dry. Crystallization of the oily residue was induced using diluted dry HCl-ether . The solid salt was then filtered off and washed with cold ether. Yield: 85-95%.

Cbz-Boc-lyzine:



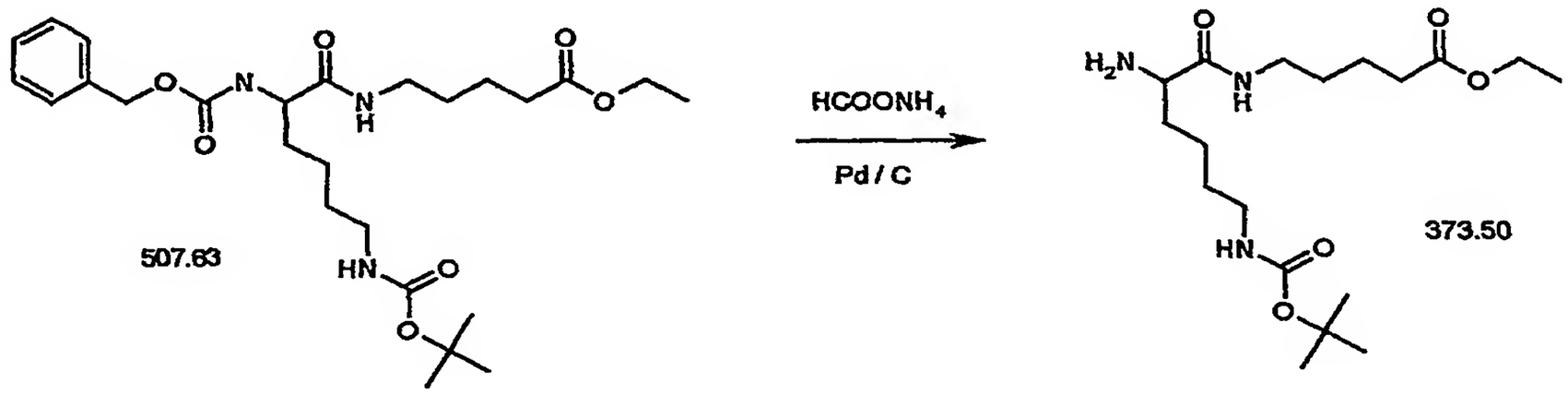
10.0 g (17.8 mmol) initial diprotected lysine dicyclohexyl-amine salt was mixed and stirred together with 2 molar equivalents of cc. H_2SO_4 and ice (ca. 40 g), while cooling with external ice water bath for 1.5-2 hours. TLC: CHCl3 : MeOH = 4:1, The $R_f \approx$ 0.7 is identical for both the starting material and the product as well, thus difference could only be detected by I_2 vaporization. The mixture was extracted 3 times with ethyl acetate. After drying, the collected organic layers were evaporated until dry, leaving a free carboxylic acid product with quasi-quantitative yield contaminated only by some solvent fraces.

Step 1: Cbz-Boc-lyzine coupling using 5-amino-valeric acid ethyl ester hydrochloride



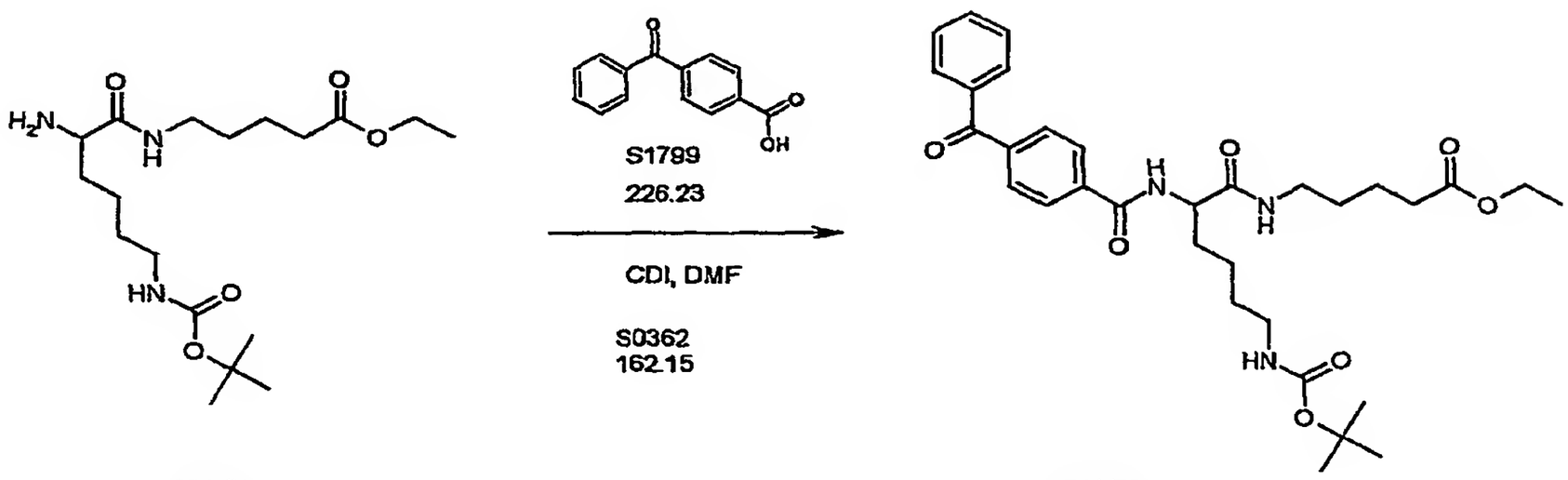
6.25 g (16.4 mmól) Cbz-Boc-lyzine was mixed with 1.05 molar equiv. of 1,1'-carbonyl-diimidazole in 1,2-dichloroethane (50 ml, HPLC grade) at room temperature and stirred for 0.5 hour at the same temperature under a moisture free atmosphere. (Mind the CO_2 gas evolution!). After this period, 2.0 molar equiv. of triethylamine and 1.0 molar equivalent of 5-amino-valeric acid ester hydrochloride was added to the activated acid and the mixture was stirred for 12 hours at room temperature. TLC: $CHCl_3$: MeOH = 4:1, $R_f \approx 0.8$. Work-up procedure: Extraction once each, first with 1% aq. citric acid solution, then with 5% aq. $NaHCO_3$ solution, and finally with distilled water. The organic layer was dried and fully evaporated in vacuum. Yield: 53-75%

Step 2: Cbz de-protection



4.9 g initial material (the result of step No. 1) was dissolved in a 1:1 mixture of ethanol-ethyl acetate (100-ml solvent-mixture). First, ammonium formate (4.0 molar equivalent) then 10% palladium on charcoal (490 mg; 10% by weight) was added to the above solution. The mixture was heated up and stirred efficiently at reflux temperature for about 4 hours. After the initial compound had disappeared, the mixture was filtered through a short celite-pad. The catalyst filtered off was washed 2-3 times with dichloromethane. All the filtrates were combined and evaporated under reduced pressure until dry. The residue was re-dissolved in dichloromethane and washed once with distilled water. After drying over $MgSO_4$, the organic phase was evaporated, leaving the de-protected title amine. TLC: 1,2-dichloroethane – ethanol = 5:1, $R_f \approx 0.2$. Yield: 60-90%.

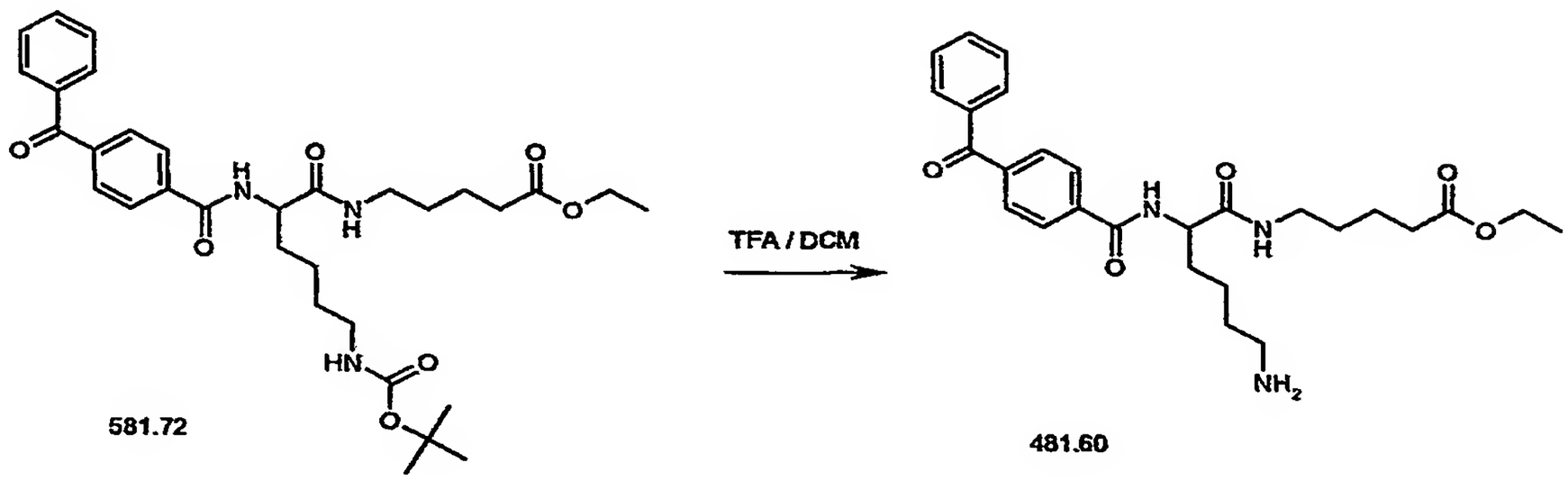
Step 3: Coupling using benzoyl-benzoic acid



373.50 581.72

1.99 g (8.8 mmol; 1.0 molar equivalent to the amine component) 4-benzoyl-benzoic acid was mixed with 1.05 molar equiv. of 1,1'-carbonyl-diimidazole in *N,N*-dimethyl-formamide (20 ml, high grade, anteriorly filtered through a silica pad) at room temperature and stirred for 1 hour at the same temperature under a moisture free atmosphere. (Mind the CO_2 gas evolution!). After this, 3.3 g (8.8 mmol) of initial amine-component (in the result of step No. 2) was added to the activated acid and the mixture was stirred for 20 hours. Work-up: The reaction mixture was fully evaporated. The residual crude product was dissolved in dichloromethane and extracted once each, first with a 1% aq. citric acid solution, then with 5% aq. $NaHCO_3$ solution, and finally with distilled water. The organic layer was dried and completely evaporated in vacuum. TLC: $CHCl_3$: MeOH = 4:1, $R_f \approx 0.8$. Yield: 76 %.

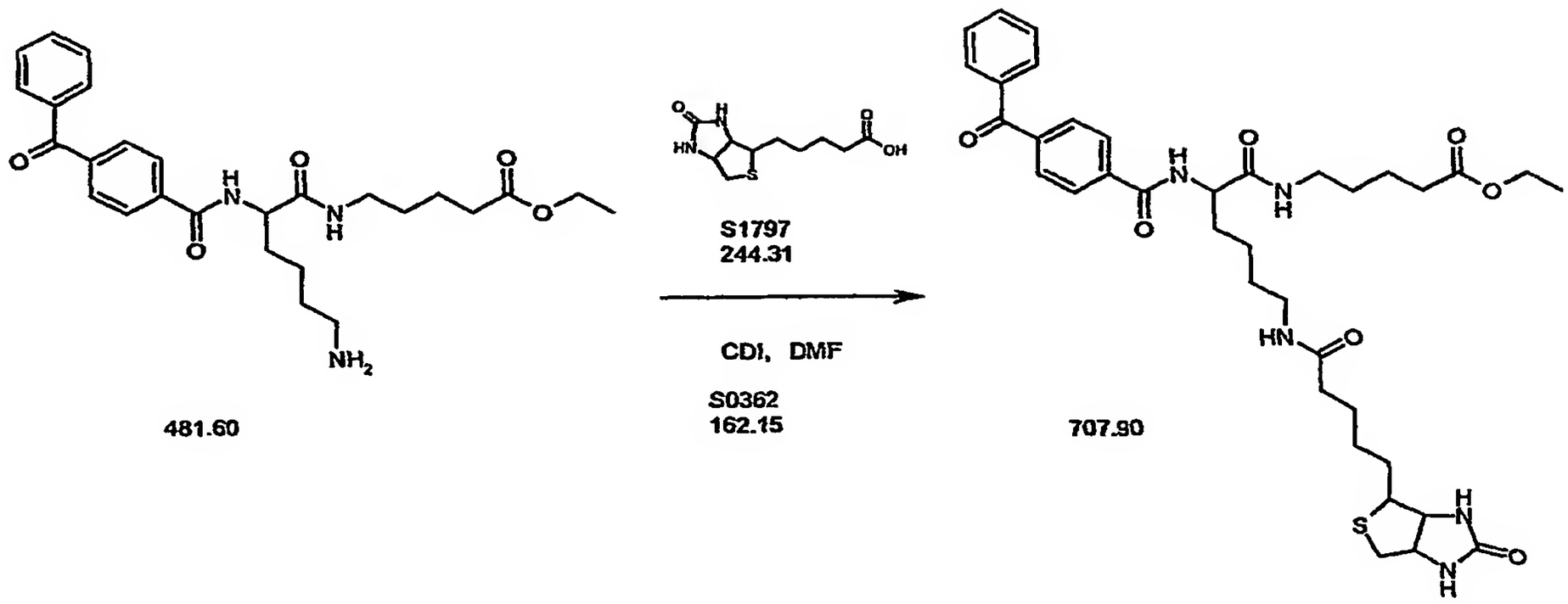
Step 4: Removal of the Boc-protecting group



The initial compound (4 g (6.9 mmol) was dissolved in a solution of trifluoroacetic acid in dichloromethane (60 ml, 30%) and stirred at room temperature until the initial compound had disappeared. The reaction was monitored by TLC, eluent: dichloroethane: ethanol = 5:1, $R_f \approx 0.2$. Work up: the solvent was removed in vacuo. The residue was dissolved in water and extracted twice with diethylether. The collected aqueous phase was basified with 5% aqueous K_2CO_3 solution to pH = 11-12 and extracted three times with chloroform. The organic phase was dried and evaporated. Yield: 85%.

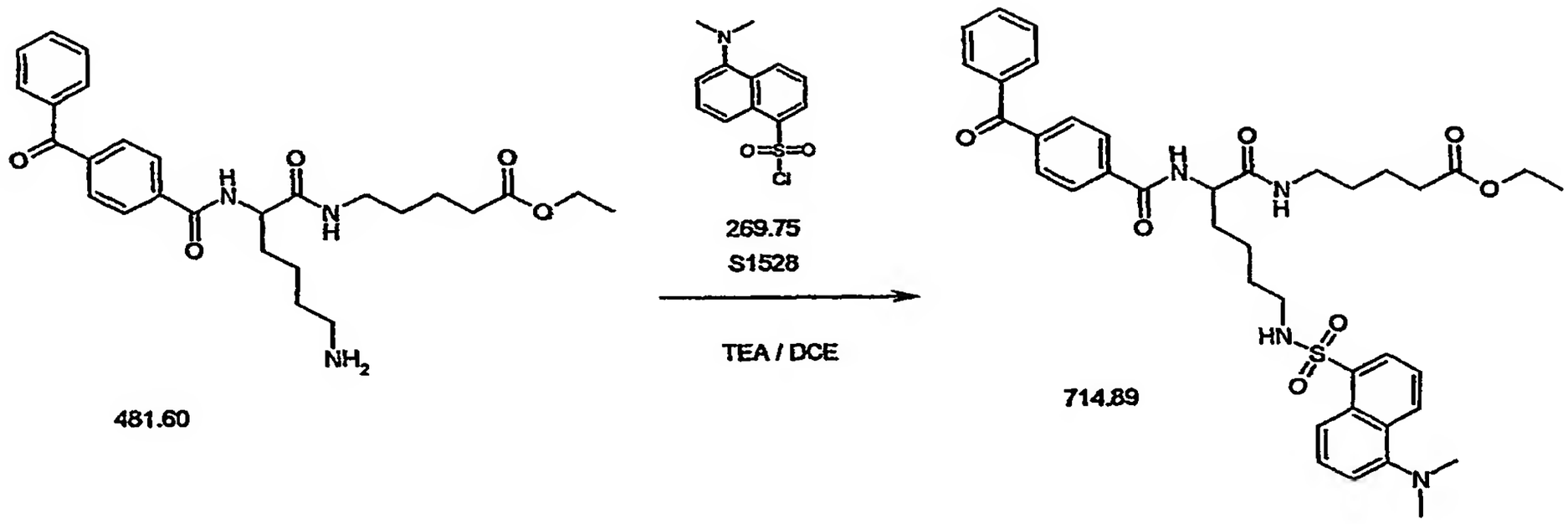
Note: The product is stable in trifluoro acetate salt form for a long time.

Step 5a: Attachment using biotin



Biotin (1.1 g, 0.8 equivalent calculated for the initial compound) was dissolved in hot *N,N*-dimethylformamide (30 ml) at 80°C, then CDI was added (0.8 equivalent calculated for the initial compound). The reaction mixture was kept at 80°C for ten minutes until the bubbling had stopped, then the heating was terminated and the mixture was stirred for another two hours. A solution consisting of the product of the previous step mixed with 30 ml of dimethylformamide was added to the above prepared reaction mixture at room temperature and stirred for another 20 hours. Work-up: the solvent was removed in vacuo, and the residue suspended in dichloromethane and filtered off. The precipitate was washed once with dichloromethane, then with a solution of citric acid (1% in water), then with solution of Na_2CO_3 (5% in water), then twice with distilled water. TLC: chloroform : methanol = 4:1, $R_f \approx 0.6$. Yield: 50%.

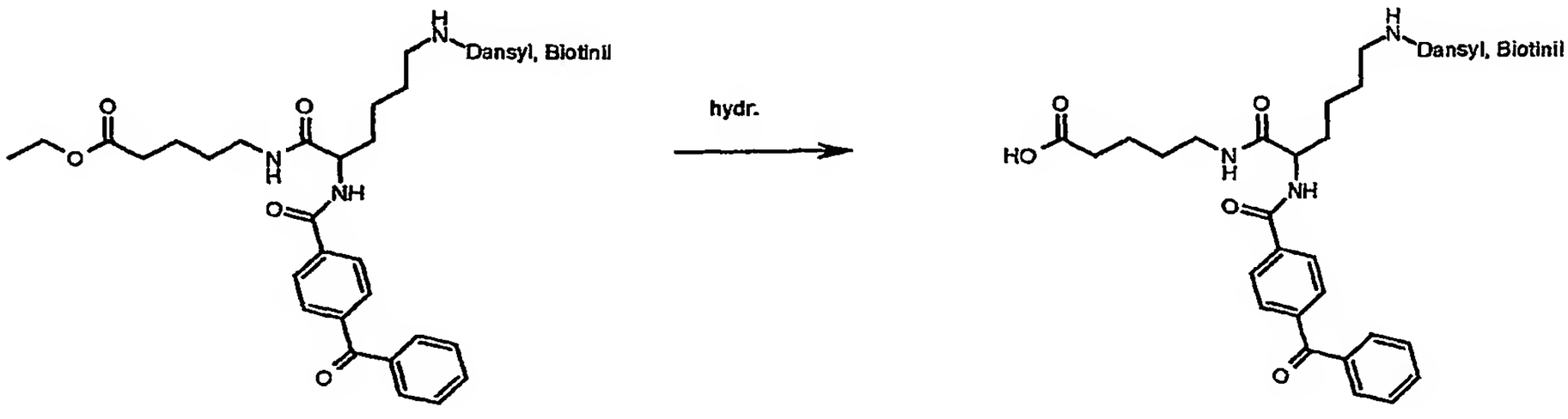
Step 5b: Coupling using dansyl chloride



To a solution of the initial compound (1.5 g, 3.1 mmol) in dichloroethane (20 ml), 2 equiv. triethylamine and 1 equiv. dansyl chloride was added and the reaction mixture stirred for hours. When the initial compound had disappeared, the mixture was extracted once with a solution of citric acid (1% in water) then twice with distilled water. The organic phase was dried then evaporated. The product was crystallized with n-hexane. TLC: hexane : ethylacetate = 10:1, R_f: 0.3. Yield: 80%.

Step 6.a. and 6.b.: Hydrolysis

The following protocol is applicable for both biotin and dansyl derivatives:



Photomarker unit

To a solution of the initial compound (1g) in ethanol (10-15 ml), 2 equivalent of aq. NaOH solution (1N) was added and the reaction mixture was stirred at room temperature until the starting compound disappeared. TLC for the dansyl substituted compound: dichloroethane: ethanol = 5:1, and for the biotin substituted compound: chloroform: methanol = 4:1. Work-up: the ethanol was removed in vacuo and the aqueous phase was acidified with an aq. HCl solution (5%). The precipitate was filtered off and washed with water. Yield: 90%.

Example 2

Coupling of amines to a biotin substituted compound

A biotin substituted compound (product of step 6a, 0.2 mmol) was dissolved in *N,N*-dimethylformamide (2-3 ml) at 80°C and 1 equivalent of CDI was added. The mixture was stirred for 1 hour, during such time the mixture cooled down to room temperature. Then 1 equivalent of a primary or secondary amine was added and the mixture was stirred at room temperature for 20 hours. The solvent was removed in vacuo, and the residue dissolved in chloroform and extracted first with solution of citric acid (1% in water), then with a solution of Na_2CO_3 (5% in water), then with distilled water. The organic phase was dried and evaporated. Further purification was performed using preparative HPLC.

Example 3.

Coupling of amines to the benzophenone or dansyl substituted compound

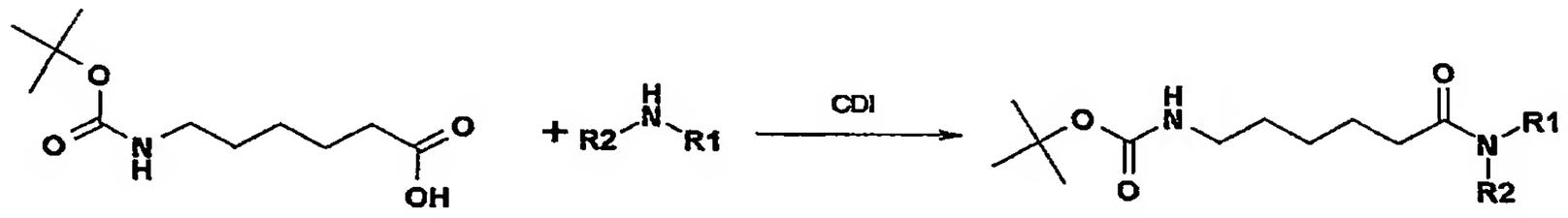
A benzophenone or dansyl substituted compound (the product of step 6b, 0.2 mmol) was dissolved in dichloroethane (2-3 ml), and 1 equivalent of CDI was added. The mixture was stirred for 1 hour at room temperature. Then 1 equivalent of primary or secondary amine was added and the mixture was stirred at room temperature for 20 hours. The reaction mixture was extracted with a solution of citric acid (1% in water), then with a solution of Na_2CO_3 (5% in water), and finally with distilled water.

The organic phase was dried and evaporated. Further purification was performed using preparative HPLC.

Example 4.

Protocols for tethering

1. Coupling of 6-(BOC-Amino) caproic acid and amines:



1 mmol, 231-mg 6-(BOC-Amino) caproic acid was dissolved in 5-ml dichloroethane. (Romil grade, SPS purity). 1.05 mmol 1,1'-carbonyldiimidazole was added to it. After the gas evolution had completed (0.5-1 h), 1.05 equiv. of amino-component was added to the reaction mixture. The progress of the reaction was monitored by TLC using dichloroethane-ethanol in a 5:1 mixture as an eluent. Iodine vapour was used to render spots visible. When the reaction had completed, the mixture was washed with 3% aqueous HCl, then with 5% aqueous Na_2CO_3, and finally with water. Product purity was checked by TLC using the eluent mentioned above. The extraction was repeated whenever necessary. The organic phase was evaporated and the residue was directly transferred to the next step.

2. Deprotection:

Boc-protected amino acid amides originating from the previous step were dissolved in 2-ml of dichloromethane. The flask was cooled in an ice water bath and 2-ml of trifluoroacetic acid was added to the cold solution dropwise. The progress of the reaction was monitored by TLC using dichloroethane-ethanol in a 5:1 mixture as an eluent. Iodine vapour was used to render spots visible. When the reaction had completed (1h approx.), the reaction mixture was evaporated. The residue was diluted with water and extracted with diethylether in order to remove the traces of the initial material. Then the aqueous phase was basified with a 20% aqueous Na_2CO_3 solution. The product was extracted with dichloromethane. The organic phase was dried over $MgSO_4$, filtered, and then the filtrate was evaporated.

CLAIMS

1. Combinatorial protein marker small molecule libraries, containing groups that may be chemically or photochemically activated, with or without reporter groups, connected at different diversity points or spatial arrangements via side chains/tethers around a common molecular core.

2. Combinatorial libraries of marker units having chemically or photochemically reactive marker groups, reporter groups, and various types of side chains attached variably to a common molecular, preferably lysine-based, structural core.

3. A library described under Claim 2, which contains benzophenone-, nitro-phenilazide groups as marker groups; biotin-, fluorescent groups as reporter groups; saturated carbon chains as side chains; and polyethylene-glycol units.

4. A method for combinatorial chemical tethering, where a side chain having a terminal functional group, preferably an amino group, and to which a marker group is optionally attached, is introduced in the optimal position to ensure structural diversity and protein binding effectiveness.

5. Application of tethered, combinatorial libraries toward the study of non-covalent interaction between affinity-based biopolymers, preferably proteins and small molecules such as ligands, substrates, and other compounds, either directly or in a form whereby they are immobilized on a solid support, preferably using affinity chromatography, chemical microarrays or microchips.

6. The application of the tethered molecules immobilized to the solid support described under Claim 5 toward the study of interaction between macromolecules, preferably proteins and small molecules, so that they are

used in established protocols or processes for using proteins, DNA, or any other type of reader chip, that have been developed for application to the study of molecules immobilized to microchips or microarrays developed other techniques.

7. A method of robotized, parallel, derivatization, in which the side chain or, directly, the marker unit are linked to the intermediate of the library.

8. The application of a high throughput biological test that enables parallel affinity labeling of a large number of samples obtained from different tissues including the detection and separation of covalently bound proteins. This method is also suitable for identifying protein markers specific to diseases and may also be used as a diagnostic method. Application of the covalent labeling method contributes to simplifying complex proteomics.

9. The application of a high-throughput analytical method suitable for sequencing the covalently bound proteins using mass spectrometric methods and for comparing them to known sequence databases.